Smart City data via LOD/LOG Service

http://mediacrawler.disit.org/dbpedia-graph/

Pierfrancesco Bellini, Paolo Nesi, Nadia Rauch

Distributed Systems and Internet Technology Lab, DISIT, http://www.disit.dinfo.unifi.it
Dipartimento Ingegneria dell'Informazione, University of Florence, Italy
pierfrancesco.bellini@unifi.it,paolo.nesi@unifi.it, nadia.rauch@unifi.it

Keywords: Smart-city services, city and mobility ontology, data alignment

Abstract

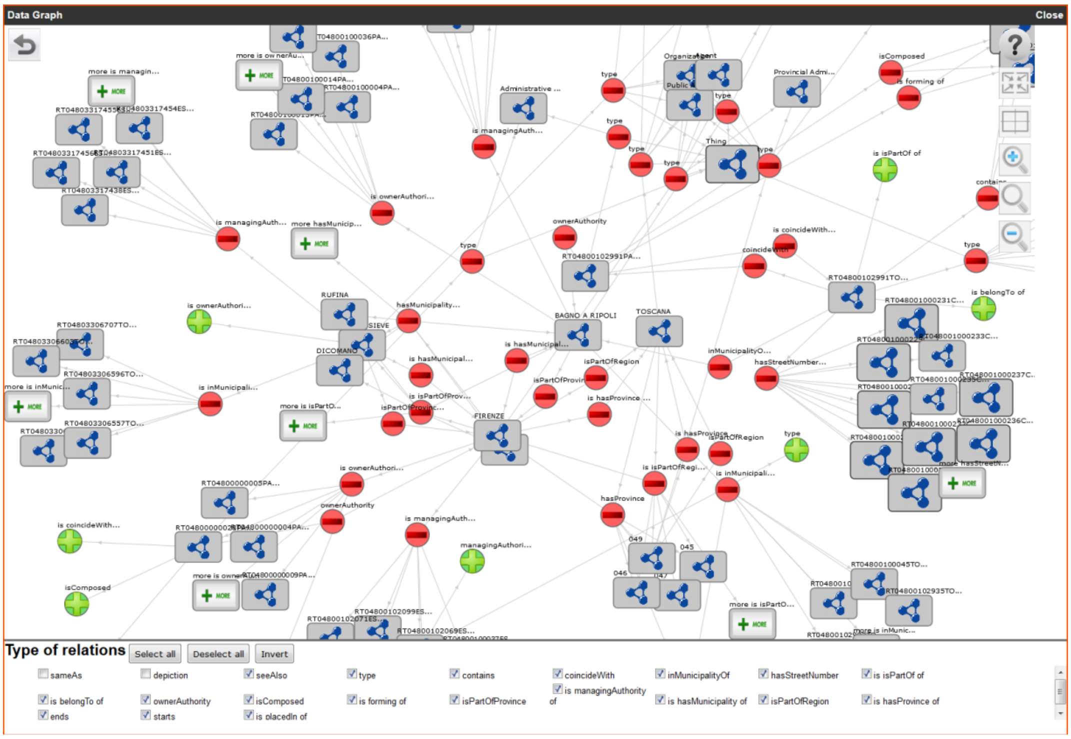
A variety of Open/Closed Data information sources are available from public administrations ranging from structural, statistical to real-time information. In most cases, this information presents inconsistencies, incompleteness, and their semantic description is not sufficient to automatically compose them to have integrated global information of the area. On the other hand, the integration of this information in a unique knowledge base, LOD and service could provide an easier and consistent access to the information to enable the development of new services. Smart city applications may take advantage to get access at integrated information including: maps, traffic status, weather conditions and forecast, parking status, pollution levels, energy, real time sensors on public and private vehicles, point of interests in the city as museums, monuments, restaurants, hotels, hospitals, etc. but also statistical data like travel accidents, per street per year. All this information, if well-structured and connected can be used to provide new services to citizens. However, the efficient management of this information it is not easy and specific tools and systems have to be developed for shortening the activation of the services in new areas and the integration of new data [1].

In this domain, the activity of DISIT lab is mainly related to a number of smart city projects among them Sii-Mobility which aims at collecting and exploiting data by solving the above mentioned problems and providing integrated data to be used for implementing smart city services for citizens mobility, public administrations, and SMEs.

As a first step, we analysed more than 400 Open Data sets from Tuscany Region, several provinces, and municipalities, data of private institutions of the area. The data have been classified in:

- *Maps and geographical information*: formed by classes Road, Node, RoadElement, AdministrativeRoad, Milestone, StreetNumber, RoadLink, Junction, Entry, and EntryRule Maneuver, is used to represent the entire road system of Tuscany region, including the permitted manoeuvres and the rules of access to the limited traffic zones;
- *Point of Interest*: economical services (public and privates), activities, which may be useful to the citizen and who may have the need to search for and to arrive at. The classification of individual services and activities will be based on the division into categories planned at regional level;
- *Transport*: data coming from major public transportation companies including scheduled times, the rail graph, and data relating to real time passage at bus stops. This groups include data of classes bus line, Ride, Route, record, RouteSection, BusStopForeast, RouteLink, and also information related to taxi, car and bike sharing;
- *Sensors*: concerning data coming from sensors installed along some city streets and surrounding areas, and those relating to free places in the main car parks of the region, have been integrated in the ontology. The sensors may include information such as pressure, humidity, pollution, car flow, car velocity, number of passed cars and tracks, etc.;
- *Weather* including status and forecasts from the consortium Lamma in Tuscany;
- *Administration*: includes information coming from public administrations as municipalities, provinces and regions which include resolutions issued by each administration, planned events, changes in the traffic arrangement, planned VIP visits, sports events, etc.;

In reality, many other details are also present and the information collected from the Open Data includes temporal aspects (dates and times, periods, time codes), citations to very important historical names (VIP names, that can be connected to dbpedia names), references and citations of geographical names (GPS, toponyms, Km from the beginning of a street, street names, geo codes, etc.), law codes and names, etc. In most cases these references contained into Open Data are inconsistent and incomplete and are the first problem to be solved in order to take advantage from the integrated data. In order to solve most of the above mentioned problems, an ingestion architecture to pass from OD to LOD has been realized.
The provided model is based on the so called *DISIT Sii-Mobility Smart-city & Mobility Ontology*. The ontology reused the following vocabularies: *OTN*: an ontology of traffic networks [2] that is more or less a direct encoding of GDF (Geographic Data Files) in OWL; *dcterms*: set of properties and classes maintained by the Dublin Core Metadata Initiative; *foaf*: dedicated to the description of the relations between people or groups; *vCard*: for a description of people and organizations; *wgs84_pos*: vocabulary representing latitude and longitude, with the WGS84 Datum, of geo-objects. The knowledge base has been built by using OWLIM-SE RDF Store instance [3].
The process also included a set of algorithms and tools that periodically collected the updated version of OD (real time data) and bring them into the LOD repository, solving inconsistencies and updating incompleteness, re-establishing links and connections, etc. To allow the regular update of data a scheduler is used to schedule the regular retrieve of data from realtime sources and the update of less dynamic data sources. In order to validate the ingestion performed a set of SPARQL queries were used. After that a validation and reconciliation activity can be done in order to join data coming from different sources as for example to join services with the road map using the street address names that are written in different ways (e.g., “Via XXVII Aprile” and “VIA VENTISETTE APRILE”) producing ‘owl:sameAs’ triples to link them; differences in encoding the name of the municipality, the lack of Istat code associated with weather forecasts; missing, wrong or incomplete street of a service (e.g., museum, pharmacy, hotel), etc.
For example in terms of size, the system ingests the data coming from the Municipality of Florence, the Tuscany Region and MIIC. Considering only files related to the daily weather forecast of all the available municipalities, we have about 286 files updated twice a day, each of which, containing also 16 lines of weather prediction for the week, we obtain an increase of approximately 270,000 HBase lines per month that, in terms of triples, corresponds to about 4 million of new triples per months.



An example of services that could be provided using the knowledge base are proximity services: like the case of a citizen who returns home by bus after work could use these services to verify if there is a supermarket near a bus stop of his busline, optimizing so his movements with his needs.
Another example could be when approaching the place of destination, a citizen at the wheel might need to check which is the car park more close to him and if there are free places inside the structure.

Moreover, in order to be able to visualize and exploit the available LOD, a tool has been developed to allow exploring semantic graph of the relations among the entities (see Figure). It is available for applications developers to explore and understand better the data available in the ontology, and can be applied to a number of LOD providing a sparql endpoint. The LOD browsing tool is called Linked Open Graph (http://mediacrawler.disit.org/dbpedia-graph/) and that can be used by PAs and SMEs accessing to many different LOD repositories to understand and test them before exploiting the LOD via API. The provided link of Linked Open Graph allows you to access at: Europeana, dbPedia, Sii-Mobility, LinkedGeoData, and CulturaItalia. You can provide with your sparql entry point and example.

References

[1] Bellini P., Di Claudio M., Nesi P., Rauch N., "Tassonomy and Review of Big Data Solutions Navigation", Big Data Computing To Be Published 26th July 2013 by Chapman and Hall/CRC
[2] Ontology of Trasportation Networks, Deliverable A1-D4, Project REWERSE, 2005 http://rewerse.net/deliverables/m18/a1-d4.pdf
[3] Barry Bishop, Atanas Kiryakov, Damyan Ognyanoff, Ivan Peikov, Zdravko Tashev, Ruslan Velkov, “OWLIM: A family of scalable semantic repositories”, Semantic Web Journal, Volume 2, Number 1 / 2011.